



Hillion, E., & Johnson, O. T. (2017). A proof of the Shepp-Olkin entropy concavity conjecture. *Bernoulli*, 23(4B), 3638-3649.  
<https://doi.org/10.3150/16-BEJ860>

Publisher's PDF, also known as Version of record

Link to published version (if available):  
[10.3150/16-BEJ860](https://doi.org/10.3150/16-BEJ860)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the final published version of the article (version of record). It first appeared online via Project Euclid at <http://projecteuclid.org/euclid.bj/1495505104>. Please refer to any applicable terms of use of the publisher.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

# A proof of the Shepp–Olkin entropy concavity conjecture

ERWAN HILLION<sup>1</sup> and OLIVER JOHNSON<sup>2</sup>

<sup>1</sup>Aix Marseille Université, CNRS, Centrale Marseille, I2M UMR 7373, 13453 Marseille, France.

E-mail: [erwan.hillion@univ-amu.fr](mailto:erwan.hillion@univ-amu.fr)

<sup>2</sup>School of Mathematics, University of Bristol, University Walk, Bristol BS8 1TW, UK.

E-mail: [o.johnson@bristol.ac.uk](mailto:o.johnson@bristol.ac.uk)

We prove the Shepp–Olkin conjecture, which states that the entropy of the sum of independent Bernoulli random variables is concave in the parameters of the individual random variables. Our proof refines an argument previously presented by the same authors, which resolved the conjecture in the monotonic case (where all the parameters are simultaneously increasing). In fact, we show that the monotonic case is the worst case, using a careful analysis of concavity properties of the derivatives of the probability mass function. We propose a generalization of Shepp and Olkin’s original conjecture, to consider Rényi and Tsallis entropies.

**Keywords:** Bernoulli sums; concavity; entropy; Poisson binomial distribution; transportation of measure

## 1. Introduction

This paper considers a conjecture of Shepp and Olkin [17], that the entropy of Bernoulli sums is a concave function of the parameters. We write  $\mathcal{B}(p)$  for the law of a Bernoulli variable with parameter  $p$ . Let  $(p_1, \dots, p_n) \in [0, 1]^n$  be a  $n$ -tuple of parameters, and consider independent random variables  $(X_1, \dots, X_n)$  with  $X_i \sim \mathcal{B}(p_i)$ . We set  $S = \sum_{i=1}^n X_i$  and, for  $k \in \{0, \dots, n\}$ , we write  $f_k := \mathbb{P}(S = k)$  for the probability mass function of  $S$ , defining a probability measure supported on  $\{0, \dots, n\}$ . Note that the law of  $S$  is sometimes referred to as a Poisson binomial distribution.

For each  $k \in \{0, \dots, n\}$ , the probabilities  $f_k$  can be seen as a smooth function of the  $n$  parameters  $\mathbf{p} := (p_1, \dots, p_n)$ . For instance, we have  $f_0 = (1 - p_1) \cdots (1 - p_n)$  and  $f_n = p_1 \cdots p_n$ . A particular case is obtained when the parameters  $p_1 = \cdots = p_n = p$  are all equal. In this case,  $(f_k)$  describes the binomial measure  $\text{Bin}(n, p)$ . This paper is focused on the study of the Shannon entropy  $H$  of  $(f_k)$ .

**Definition 1.1.** Writing  $U(x) := x \log(x)$  if  $x > 0$  and  $U(0) := 0$ , we define:

$$H(\mathbf{p}) := - \sum_{k=0}^n U(f_k) := - \sum_{k=0}^n f_k \log(f_k). \quad (1.1)$$

The entropy  $H$  can itself be seen as a smooth function of the parameters  $p_1, \dots, p_n$ . This article is devoted to the proof of the following.

**Theorem 1.2 (Shepp–Olkin theorem).** *For any  $n \geq 1$ , the function  $\mathbf{p} \mapsto H(\mathbf{p})$  is concave.*

We simplify notation somewhat by considering the case where each  $p_i := p_i(t)$  is an affine function of parameter  $t \in [0, 1]$ , so that the derivative  $p'_i := \frac{d}{dt} p_i(t)$  is constant in  $t$ . Theorem 1.2 will follow if we can show that the entropy is concave in  $t$ .

Theorem 1.2 was conjectured by Shepp and Olkin [17]. In their original paper, Shepp and Olkin stated that the conjecture is true in the cases  $n = 2$  and  $n = 3$ , and proved that it holds for the binomial case where all  $p_i$  are identical (see also Mateev [13]). Since then, progress has been limited. In [23], Yu and Johnson considered the thinning operation of Rényi [15], and proved a result which implies concavity of entropy when each  $p_i(t)$  is proportional to  $t$  or  $1 - t$ . Further, [10], Theorem 1.1, proved Theorem 1.2 in the case where each  $p_i(t)$  is either constant or equal to  $t$ .

More significant progress was made in [11] by the present authors, who proved that the entropy is concave when all  $p'_i$  have the same sign. Perhaps surprisingly, in the current paper we show that this ‘monotone’ case resolved in [11] is the most difficult.

The study of the entropy in the monotone case was motivated by the theory of transportation of discrete probability measures and the introduction of a distance defined by a discrete form of the Benamou–Brenier formula [3] (see also [9]). This idea of a discrete geodesic was designed as an analogue to the more developed theory of geodesics on continuous spaces such as Riemannian manifolds (see, for example, [21]), where concavity of the entropy relates to the Ricci curvature of the underlying manifold (see [6, 12, 18, 19]). Further, using ideas related to the Bakry–Émery  $\Gamma$ -calculus [2], concavity of the entropy can be used to prove functional inequalities, such as log-Sobolev or HWI (see, for example, [1, 7]). It remains an important problem to provide discrete analogues of this theory.

The structure of the paper is as follows: in Section 2 we state some technical results required for the proof of Theorem 1.2, the proofs of which are deferred to the Appendix. In Section 3, we prove Theorem 1.2 itself. In Section 4, we propose a generalized form of Shepp and Olkin’s conjecture, in terms of Rényi and Tsallis entropies  $H_{R,q}$  and  $H_{T,q}$ .

## 2. Technical results required in the proof

The strategy of the proof of the ‘monotone’ form of the Shepp–Olkin conjecture in [11] was to show that the entropy is a concave function along an interpolating (geodesic) path between discrete random variables, assuming certain conditions are satisfied. In particular, [11], Theorem 4.4, showed that if a ‘ $k$ -monotonicity’ condition, a generalized log-concavity condition, and so-called Condition 4 hold, then the entropy is concave. At the heart of this analysis was control of terms of the form  $U(1 - x)$  by the second-order Taylor expansion of  $U$ . This relied on the fact that a term we refer to as  $B_k$  (see (3.6)) satisfies  $B_k \leq 0$ .

However, unfortunately the inequality  $B_k \leq 0$  does not hold in general, so in this paper we will use Lemma 2.1 to control terms of this kind. The strategy is essentially to show that Condition 4 (here referred to as ‘bound on  $h$ ’, Proposition 2.4) holds in the general Shepp–Olkin case. Using this result, we deduce Corollary 2.5 which plays the part of the  $k$ -monotonicity and generalized

log-concavity conditions. For the sake of simplicity, we restrict our discussion to mass functions which are Bernoulli sums, but a version of Theorem 1.2 will hold for interpolating paths made up of log-concave  $f$  for which the bound on  $h$ , Proposition 2.4, holds.

We now state a number of technical results which are required in the proof of Theorem 1.2, the main result of the paper. The proofs are deferred to the [Appendix](#).

## 2.1. Concavity of functions

We first state a technical result concerning certain functions  $U$ :

**Lemma 2.1.** *Let  $U : (0, \infty) \mapsto \mathbb{R}$  be a function such that (i)  $U(1) = 0$ , (ii)  $U'(1) = 1$ , (iii)  $U'''(t) \leq 0$  for all  $t$  and (iv)  $\log U''(t)$  is convex in  $t$ .*

*For  $A, B, C, \alpha, \beta, \gamma$  satisfying  $0 < A < 1, 0 < C < 1$  and  $B^2 \leq AC, \beta^2 \leq \alpha\gamma$  we have:*

$$\alpha U(1 - A) - 2\beta U(1 - B) + \gamma U(1 - C) \geq -\alpha A + 2\beta B - \gamma C. \quad (2.1)$$

Note that the conditions of this lemma are satisfied for  $U(x) = x \log x$ .

## 2.2. Cubic inequality for Bernoulli sums

**Proposition 2.2.** *Let  $T = X_1 + \dots + X_m$  be the sum of independent Bernoulli variables, with probability mass function  $\pi_k := \mathbb{P}(T = k)$ , then:*

$$\pi_{k-2}\pi_{k+1}^2 + \pi_k^3 + \pi_{k-1}^2\pi_{k+2} \geq \pi_{k-2}\pi_k\pi_{k+2} + 2\pi_{k-1}\pi_k\pi_{k+1} \quad \text{for any } k. \quad (2.2)$$

Note that for  $k$  outside the support of  $S$ , each term is equal to zero, so the inequality is trivially true. In the [Appendix](#), we show that Proposition 2.2 can be proved directly, using two other cubic inequalities (A.7) and (A.8) taken from [11].

If we write  $D_k := \mathcal{L}(\pi_k) = \pi_k^2 - \pi_{k-1}\pi_{k+1}$ , then it is well known (see Proposition A.1) that  $D_k \geq 0$ , a result referred to as log-concavity of  $\pi_k$ . Observe that (2.2) is equivalent to the statement that  $\mathcal{L}^2(\pi_k) = \mathcal{L}(D_k) = D_k^2 - D_{k-1}D_{k+1} \geq 0$ , that the  $D_k$  themselves are log-concave, a property is referred to as 2-fold log-concavity.

This result also follows from a result of Brändén [5], which resolved a conjecture made independently by Stanley, by McNamara and Sagan and by Fisk (see Brändén's paper [5] for details). Brändén discusses conditions under which infinite log-concavity (the fact that the iterated  $\mathcal{L}^r(\pi_k) \geq 0$  for all  $r \geq 1$ ) holds.

## 2.3. Upper bounding $h_k$

Recall that we consider the random variable  $S = \sum_{l=1}^n X_l$  with probability mass function  $f_k(t)$ . For  $i \in \{1, \dots, n\}$  we define  $S^{(i)} := \sum_{l \neq i} X_l$ , with probability mass function  $f_k^{(i)}$  supported on

$\{0, \dots, n-1\}$ . Similarly, for a pair of indices  $i \neq j \in \{1, \dots, n\}$  we define  $S^{(i,j)} := \sum_{l \notin \{i,j\}} X_l$  with mass function  $f_k^{(i,j)}$  supported on  $\{0, \dots, n-2\}$ .

**Definition 2.3.** As in [11], we make the following definitions:

$$g_k := \sum_i p'_i f_k^{(i)} := \sum_i p'_i \mathbb{P}(S^{(i)} = k), \quad (2.3)$$

$$h_k := \sum_{i \neq j} p'_i p'_j f_k^{(i,j)} := \sum_{i \neq j} p'_i p'_j \mathbb{P}(S^{(i,j)} = k). \quad (2.4)$$

We prove a strong upper bound on  $h_k$ , which lies at the heart of the proof of Theorem 1.2.

**Proposition 2.4 (Bound on  $h$ ).** For  $k = 0, \dots, n-2$ :

$$h_k(f_{k+1}^2 - f_k f_{k+2}) \leq 2g_k g_{k+1} f_{k+1} - g_k^2 f_{k+2} - g_{k+1}^2 f_k. \quad (2.5)$$

**Corollary 2.5.** For  $k = 0, \dots, n-2$ , the  $h_k f_k \leq g_k^2$  and  $h_k f_{k+2} \leq g_{k+1}^2$ .

In [11], Proposition 2.4 is referred to as Condition 4, and is proved for the case where all  $p'_i$  have the same sign. In fact, we show using Proposition 2.2 that this inequality holds in general.

### 3. Proof of the Shepp–Olkin conjecture

Theorem 1.2 is obvious if  $n = 1$ . We now fix some  $n \geq 2$  and take  $X_1, \dots, X_n$  to be independent Bernoulli variables with parameters  $p_1, \dots, p_n$ , where each  $p_i = p_i(t)$  is an affine function of  $t$  (with constant derivative  $p'_i$ ).

**Proposition 3.1.** To prove Theorem 1.2, it suffices to show that  $u_k \geq 0$  for any choice of parameters  $(p_1, \dots, p_n)$ , of slopes  $(p'_1, \dots, p'_n)$  and index  $k \in \{0, \dots, n-2\}$ , where we write:

$$u_k := h_k \log\left(\frac{f_k f_{k+2}}{f_{k+1}^2}\right) + \left(\frac{g_k^2}{f_k} - 2\frac{g_k g_{k+1}}{f_{k+1}} + \frac{g_{k+1}^2}{f_{k+2}}\right). \quad (3.1)$$

**Proof.** As in [11], direct calculation (for example, using the probability generating function) shows that the first two derivatives of  $f_k$  satisfy

$$\frac{df_k}{dt}(t) = g_{k-1} - g_k, \quad (3.2)$$

$$\frac{d^2 f_k}{dt^2}(t) = h_k - 2h_{k-1} + h_{k-2}. \quad (3.3)$$

Hence, we can write the derivative of the entropy:

$$\begin{aligned}
 H''(t) &= - \sum_{k=0}^n \frac{d^2}{dt^2} U(f_k(t)) \\
 &= - \sum_{k=0}^n U''(f_k) \left( \frac{df_k}{dt} \right)^2 - \sum_{k=0}^n U'(f_k) \left( \frac{d^2 f_k}{dt^2} \right) \\
 &= - \sum_{k=0}^n U''(f_k) (g_{k-1} - g_k)^2 - \sum_{k=0}^n U'(f_k) (h_k - 2h_{k-1} + h_{k-2}) \quad (3.4) \\
 &\leq - \sum_{k=0}^{n-2} [(g_k^2 U''(f_k) - 2g_k g_{k+1} U''(f_{k+1}) + g_{k+1}^2 U''(f_{k+2})) \\
 &\quad + h_k (U'(f_k) - 2U'(f_{k+1}) + U'(f_{k+2}))]. \quad (3.5)
 \end{aligned}$$

The form of equation (3.4) follows using (3.2) and (3.3). The relabelling in (3.5) uses the fact that  $g_k$  is supported on  $\{0, \dots, n-1\}$  and  $h_k$  supported on  $\{0, \dots, n-2\}$ . This expression is an inequality since (taking into account the end points of the range of summation) we remove the terms  $-g_{n-1}^2 U''(f_{n-1}) - g_0^2 U''(f_1)$ , which are negative assuming that  $U'' \geq 0$ .

Making the choice of  $U(x) = x \log x$  we deduce the form of  $u_k$  given in (3.1), since  $U''(x) = 1/x \geq 0$  in this case.  $\square$

One of the main differences with the monotonic case studied in [11] is that the quantities  $(g_l)$  and  $(h_l)$  are not necessarily positive. However, we note the following lemma.

**Lemma 3.2.** *If  $h_k \leq 0$ , then  $u_k \geq 0$ .*

**Proof.** This follows quite easily from the log-concavity property for the  $(f_l)$ , see Proposition A.1, which implies that  $h_k \log(\frac{f_k f_{k+2}}{f_{k+1}^2}) \geq 0$ . Further, using the arithmetic-geometric inequality and then again the log-concavity property, we have:

$$\frac{g_k^2}{f_k} + \frac{g_{k+1}^2}{f_{k+2}} \geq 2 \frac{|g_k| |g_{k+1}|}{\sqrt{f_k f_{k+2}}} \geq 2 \frac{|g_k| |g_{k+1}|}{f_{k+1}} \geq 2 \frac{g_k g_{k+1}}{f_{k+1}},$$

which proves that  $u_k \geq 0$ .  $\square$

Hence, from now on, we can suppose that  $h_k > 0$ .

**Proof of Theorem 1.2.** We set:

$$A_k := \frac{g_k^2 - f_k h_k}{g_k^2}, \quad B_k := \frac{|g_k| |g_{k+1}| - f_{k+1} h_k}{|g_k| |g_{k+1}|}, \quad C_k := \frac{g_{k+1}^2 - f_{k+2} h_k}{g_{k+1}^2}, \quad (3.6)$$

and

$$\alpha_k := \frac{g_k^2}{f_k}, \quad \beta_k := \frac{|g_k||g_{k+1}|}{f_{k+1}}, \quad \gamma_k := \frac{g_{k+1}^2}{f_{k+2}}. \quad (3.7)$$

Noticing that  $\log(g_k^2) + \log(g_{k+1}^2) - 2\log(|g_k||g_{k+1}|) = 0$ , we have:

$$\log\left(\frac{f_k f_{k+2}}{f_{k+1}^2}\right) = \log\left(\frac{h_k f_k}{g_k^2}\right) - 2\log\left(\frac{h_k f_{k+1}}{|g_k||g_{k+1}|}\right) + \log\left(\frac{h_k f_{k+2}}{g_{k+1}^2}\right). \quad (3.8)$$

Straightforward calculations give:

$$u_k \geq \alpha_k U(1 - A_k) - 2\beta_k U(1 - B_k) + \gamma_k U(1 - C_k) + (\alpha_k A_k - 2\beta_k B_k + \gamma_k C_k). \quad (3.9)$$

We need to verify the assumptions of Lemma 2.1 on  $A_k$ ,  $B_k$ ,  $C_k$ . First, we can show that  $B_k^2 \leq A_k C_k$ , which is equivalent to

$$h_k(f_{k+1}^2 - f_k f_{k+2}) \leq 2|g_k||g_{k+1}|f_{k+1} - g_k^2 f_{k+2} - g_{k+1}^2 f_k. \quad (3.10)$$

However, we notice that this is implied by the stronger inequality (2.5) from the bound on  $h$ , Proposition 2.4. Second, we need to check the inequalities  $A_k \geq 0$  and  $C_k \geq 0$ . As we have proved that  $0 \leq B_k^2 \leq A_k C_k$ , it suffices to verify that  $A_k \geq 0$ , which is a restatement of Corollary 2.5.

Clearly, the log-concavity of  $f$  shows that  $\beta_k^2 \leq \alpha_k \gamma_k$  for the quantities defined in (3.7). We can thus apply Lemma 2.1 to deduce that  $u_k \geq 0$  and prove Theorem 1.2.  $\square$

## 4. Tsallis and Rényi entropy

Having resolved the Shepp–Olkin conjecture for (Shannon) entropy, it is natural to want to generalize our result to a wider class of entropy-like functionals. Recall the following definitions, each of which reduce to the Shannon entropy (1.1) as  $q \rightarrow 1$ .

**Definition 4.1.** Given a probability mass function  $f$  supported on  $\{0, \dots, n\}$ , for  $0 \leq q \leq \infty$  define

$$1. \text{ } q\text{-Rényi entropy (see [16]): } H_{R,q}(f) = \frac{1}{1-q} \log\left(\sum_{x=0}^n f_x^q\right), \quad (4.1)$$

$$2. \text{ } q\text{-Tsallis entropy (see [20]): } H_{T,q}(f) = \frac{1}{q-1} \left(1 - \sum_{x=0}^n f_x^q\right). \quad (4.2)$$

We know that  $H_{R,\infty}(f) = -\log \max_x f(x)$  (the min-entropy), which is not concave for Bernoulli sums (for example, for Bernoulli( $p$ ) with  $p < 1/2$ , this is just  $-\log(1-p)$ , which is convex in  $p$ ). For  $q = 0$ , Rényi entropy is the log of the size of the support, which is constant (and hence concave) for Bernoulli sums. This suggests the following conjecture.

**Conjecture 4.2 (Generalized Shepp–Olkin conjecture).**

1. There is a critical  $q_R^*$  such that the  $q$ -Rényi entropy of all Bernoulli sums is concave for  $q \leq q_R^*$ , and the entropy of some interpolation is convex for  $q > q_R^*$ .
2. There is a critical  $q_T^*$  such that the  $q$ -Tsallis entropy of all Bernoulli sums is concave for  $q \leq q_T^*$ , and the entropy of some interpolation is convex for  $q > q_T^*$ .

Indeed (based on Lemma 4.3 below) we conjecture that  $q_R^* = 2$  and  $q_T^* = 3.65986\dots$ , the root of  $2 - 4q + 2^q = 0$ .

We mention some limited progress towards this conjecture.

**Lemma 4.3.**

1. For any  $q > 2$ , there exists a Shepp–Olkin interpolation with convex  $q$ -Rényi entropy.
2. For any  $q > q^* := 3.65986\dots$ , there exists a Shepp–Olkin interpolation with convex  $q$ -Tsallis entropy.

**Proof.**

1. Consider the Bernoulli  $\mathcal{B}(p)$  family, for which  $T(p) = p^q + (1-p)^q$ . As  $p \rightarrow 0$ , since  $q > 2$ , the  $T(p) \rightarrow 1$ . Similarly,  $T(p)' = q(p^{q-1} + (1-p)^{q-1}) \rightarrow q$ , and  $T(p)'' = q(q-1)(p^{q-2} + (1-p)^{q-2}) \rightarrow q(q-1)$ . In equation (4.3), we obtain  $-q - \frac{q^2}{1-q} = q/(q-1) > 0$ .
2. Consider the Binomial  $(2, p)$  family, for which  $T(p) = (p^2)^q + (2p(1-p))^q + (1-p)^{2q}$ . The second derivative of  $H_{T,q}$  at  $p = 1/2$  is  $2^{3-2q}(2-4q+2^q)q/(q-1)$ , which is positive for  $q > q^*$ .  $\square$

Although Rényi and Tsallis entropies are monotone functions of one another, and so are maximised by the same mass function, the relationship between their concavity properties involves the chain rule.

**Lemma 4.4.**

1. For  $q < 1$ , if the Tsallis entropy is concave, then so is the Rényi entropy.
2. For  $q > 1$ , if the Rényi entropy is concave, then so is the Tsallis entropy.

**Proof.** If we write  $T(t) = \sum_x f_x(t)^q$  for some path, then

$$H''_{R,q}(t) = \frac{T''(t)}{(1-q)T(t)} - \frac{1}{1-q} \left( \frac{T'(t)}{T(t)} \right)^2 = \frac{1}{T(t)} H''_{T,q}(f(t)) - \frac{1}{1-q} \left( \frac{T'(t)}{T(t)} \right)^2. \quad (4.3)$$

Since the difference has a sign we can control, we can deduce the result holds.  $\square$



**Remark 4.5.** We can consider the Tsallis entropy in the framework used earlier. Using equation (3.5), we deduce that  $H''_{T,q}(t) = -q \sum_{k=0}^{n-2} u_k$ , where

$$u_k := -\frac{1}{1-q} h_k(f_k^{q-1} - 2f_{k+1}^{q-1} + f_{k+2}^{q-1}) + (g_k^2 f_k^{q-2} - 2g_k g_{k+1} f_{k+1}^{q-2} + g_{k+1}^2 f_{k+2}^{q-2}).$$

Conjecture 4.2 would follow if  $u_k \geq 0$ , or even if  $\tilde{u}_k := u_k + \nabla_1(v_k) \geq 0$ , where  $v_k$  is some function and  $\nabla_1$  represents the left discrete derivative.

There is an heuristic argument which supports such a conjecture, at least in the monotonic case. Using the fact that

$$(g_k^2 f_k^{q-2} - 2g_k g_{k+1} f_{k+1}^{q-2} + g_{k+1}^2 f_{k+2}^{q-2}) = (g_k - g_{k+1})^2 f_{k+1}^{q-2} + \nabla_1(g_{k+1}^2 \nabla_1(f_{k+2}^{q-2})),$$

by taking  $v_k = -\frac{1-q}{2-q} g_{k+1}^2 \nabla_1(f_{k+2}^{q-2})$  we can write:

$$\tilde{u}_k = -\frac{1}{1-q} h_k \nabla_2(f_{k+2}^{q-1}) + (\nabla_1 g_{k+1})^2 f_{k+1}^{q-2} + \frac{1}{2-q} \nabla_1(g_{k+1}^2 \nabla_1(f_{k+2}^{q-2})), \quad (4.4)$$

where  $\nabla_1$  and  $\nabla_2$  stand for the first and second left discrete derivatives. A possible continuous analogy consists in considering, at least at a formal level, the expression

$$u := -\frac{1}{1-q} h(f^{q-1})'' + (g')^2 f^{q-2} + \frac{1}{2-q} (g^2 (f^{q-2}))', \quad (4.5)$$

for a triple  $(f, g, h)$  of real functions. If we make the further assumption that  $g = vf$  and  $h = v^2 f$  for some velocity function  $v$ , which was already an assumption made in the heuristic study of the monotonic case of the Shepp–Olkin conjecture [11], Section 2, and referred to as a Benamou–Brenier condition, equation (4.5) is simplified into

$$u = v'^2 f^q,$$

which is clearly non-negative, and corresponds to [11], Corollary 2.6.

**Remark 4.6.** Recent work of Bobkov and Chistyakov [4] and of Wang and Madiman [22] has studied formulations of Shannon’s Entropy Power Inequality (EPI) for Rényi entropy of the convolution of probability densities  $f_i$  in  $\mathbb{R}^d$ . To be specific, [4], Theorem 1, shows that for  $q > 1$ , the Rényi entropy power  $N_{R,q}$  satisfies

$$N_{R,q}(f_1 \star f_2 \star \cdots \star f_n) \geq c_q \sum_{k=1}^n N_{R,q}(f_k), \quad (4.6)$$

for an explicit constant  $c_q$  depending on  $q$ . In contrast, [22], Main Theorem 1, shows that for any  $q$  the entropy of the convolution

$$H_{R,q}(f_1 \star f_2 \star \cdots \star f_n) \geq H_{R,q}(f_1^* \star f_2^* \star \cdots \star f_n^*), \quad (4.7)$$

where  $f_i^*$  is the spherically symmetric rearrangement of  $f_i$ .

Since a strengthened form of the EPI (originally due to Costa [8]) proves the entropy power is concave on convolution with a Gaussian, the Shepp–Olkin conjecture may relate to some form of a discrete EPI. It is interesting that [22], Section 7, proves their main theorem only using majorization theory; Shepp and Olkin's original paper [17] showed that the entropy of Bernoulli sums satisfies the (weaker) property of Schur concavity.

## Appendix: Proof of technical results

### A.1. Proof of functional inequality, Lemma 2.1

**Proof of Lemma 2.1.** We consider the function

$$\xi(t) := \alpha U(1-tA) - 2\beta U(1-tB) + \gamma U(1-tC), \quad (\text{A.1})$$

well-defined and smooth for  $0 \leq t \leq 1$ . Its derivative is given by

$$\xi'(t) = -\alpha A U'(1-tA) + 2\beta B U'(1-tB) - C \gamma U'(1-tC), \quad (\text{A.2})$$

so Inequality (2.1) can be rewritten  $\xi(1) - \xi(0) \geq \xi'(0)$ . Since, by the mean value theorem,  $\xi(1) - \xi(0) = \xi'(s)$  for some  $0 \leq s \leq 1$ , it is sufficient to show the convexity of  $\xi$ , that is, that  $\xi''(t) \geq 0$  for any  $0 < t < 1$ . We have:

$$\xi''(t) = \alpha A^2 U''(1-tA) - 2\beta B^2 U''(1-tB) + \gamma C^2 U''(1-tC) \quad (\text{A.3})$$

$$\geq 2\sqrt{\alpha\gamma A^2 C^2 U''(1-tA)U''(1-tC)} - 2\beta B^2 U''(1-tB) \quad (\text{A.4})$$

$$\geq 2\beta B^2 (\sqrt{U''(1-tA)U''(1-tC)} - U''(1-tB)) \quad (\text{A.5})$$

$$\geq 2\beta B^2 (\sqrt{U''(1-tA)U''(1-tC)} - U''(1-t(A+C)/2)) \quad (\text{A.6})$$

here (A.4) follows by the arithmetic mean-geometric mean inequality, and (A.5) follows by the assumptions  $\beta^2 \leq \alpha\gamma$  and  $B^2 \leq AC$ , and (A.6) uses the fact that by assumption (iii)  $U''(s)$  is decreasing in  $s$ .

The result follows since we can deduce the positivity of (A.6) using assumption (iv) (the log-convexity of  $U''$ )  $\square$

### A.2. Proof of Proposition 2.2

For  $\pi_k$  the probability mass function of the sum  $T$ , we note that the quadratic Newton inequality (see, for example, Niculescu [14]) gives the log-concavity of  $\pi$ .

**Proposition A.1.** *For any  $k \in \{0, \dots, m-2\}$  we have  $\pi_{k+1}^2 \geq \pi_k \pi_{k+2}$ .*

Further properties of Bernoulli sums have been proven by the authors in [11]. The most interesting for our purposes are the inequalities stated there as  $C_1(k) \geq 0$  and  $\overline{C}_1(k) \geq 0$ , which give that:

$$(\pi_{k-1}^2 - \pi_{k-2}\pi_k)\pi_{k+1} \leq \pi_{k-1}(\pi_k^2 - \pi_{k-1}\pi_{k+1}), \quad (\text{A.7})$$

$$(\pi_{k+1}^2 - \pi_k\pi_{k+2})\pi_{k-1} \leq \pi_{k+1}(\pi_k^2 - \pi_{k-1}\pi_{k+1}). \quad (\text{A.8})$$

These results have been stated (and proven by induction on the number  $n$  of parameters) as [11], Theorem A2, Corollary A3.

Multiplying equations (A.7) and (A.8) together, and rearranging, we simply obtain the positivity of

$$\pi_{k-1}\pi_k\pi_{k+1}(\pi_{k-2}\pi_{k+1}^2 + \pi_k^3 + \pi_{k-1}^2\pi_{k+2} - \pi_{k-2}\pi_k\pi_{k+2} - 2\pi_{k-1}\pi_k\pi_{k+1}),$$

and we deduce Proposition 2.2 holds.

### A.3. Proof of bound on $h$ , Proposition 2.4

In the monotonic case studied in [11], equation (2.5) was referred to as Condition 4, and was verified under the assumption that the  $p'_i$  are non-negative. More precisely, some involved manipulations (using the definitions of  $g_k$  and  $h_k$  given in equations (2.3) and (2.4)) allow us to deduce that (see Proposition 6.1 and equation (73) of [11]):

**Lemma A.2.** *We can write the term*

$$\begin{aligned} & (2g_k g_{k+1} f_{k+1} - g_k^2 f_{k+2} - g_{k+1}^2 f_k) - h_k(f_{k+1}^2 - f_k f_{k+2}) \\ &= \sum_{i < j} (p_i'^2 p_j (1 - p_j) b_{i,j} + p_j'^2 p_i (1 - p_i) b_{j,i} \\ & \quad + 2p_i' p_j' p_i (1 - p_i) p_j (1 - p_j) c_{i,j}), \end{aligned} \quad (\text{A.9})$$

where the coefficients satisfy

$$\begin{aligned} c_{i,j} := & -(f_k^{(i,j)})^3 + 2f_{k-1}^{(i,j)} f_k^{(i,j)} f_{k+1}^{(i,j)} - (f_{k+1}^{(i,j)})^2 f_{k-2}^{(i,j)} \\ & - (f_{k-1}^{(i,j)})^2 f_{k+2}^{(i,j)} + f_{k-2}^{(i,j)} f_k^{(i,j)} f_{k+2}^{(i,j)}. \end{aligned} \quad (\text{A.10})$$

Lemma 6.2 of [11] shows that

$$b_{i,j} \geq -\frac{1}{2}(p_i(1 - p_j) + p_j(1 - p_i))c_{i,j}. \quad (\text{A.11})$$

This result can be verified using the expression for  $b_{i,j}$  given in [11], equation (72), using equations (A.7) and (A.8) and other related cubic inequalities for  $\pi$ .

We now observe that  $c_{i,j} \leq 0$  by Proposition 2.2, simply by taking  $\pi_k = f_k^{(i,j)}$  in equation (2.2). Combining this with (A.11), we deduce that  $b_{i,j}$  is positive, and that (treated as a quadratic in  $p'_i$  and  $p'_j$ ), the bracketed term in equation (A.9) has negative discriminant

$$\begin{aligned} & 4p_i(1-p_i)p_j(1-p_j)c_{i,j}^2 - 4b_{i,j}b_{j,i} \\ & \leq c_{i,j}^2(4p_i(1-p_i)p_j(1-p_j) - (p_i(1-p_j) + p_j(1-p_i))^2) \\ & = -c_{i,j}^2(p_i - p_j)^2, \end{aligned}$$

meaning that it is positive for all values of  $p'_i$  and  $p'_j$ .

Note that the negativity of  $c_{i,j}$  shows that if we fix  $|p'_i|$  and  $|p'_j|$  then (A.9) is minimized when  $p'_i$  and  $p'_j$  have the same sign, justifying the claim that the monotonic case is the worst case.

#### A.4. Proof of Corollary 2.5

**Proof of Corollary 2.5.** Proposition 2.4 gives an upper bound on  $h_k$ :

$$h_k(f_{k+1}^2 - f_k f_{k+2}) \leq 2g_k g_{k+1} f_{k+1} - g_k^2 f_{k+2} - g_{k+1}^2 f_k. \quad (\text{A.12})$$

In order to prove that  $f_k h_k \leq g_k^2$ , it thus suffices to show that

$$(2g_k g_{k+1} f_{k+1} - g_k^2 f_{k+2} - g_{k+1}^2 f_k) f_k \leq (f_{k+1}^2 - f_k f_{k+2}) g_k^2. \quad (\text{A.13})$$

But this equation can be simplified into

$$(f_{k+1} g_k - f_k g_{k+1})^2 \geq 0, \quad (\text{A.14})$$

which is obviously true.  $\square$

## References

- [1] Ané, C., Blachère, S., Chafaï, D., Fougères, P., Gentil, I., Malrieu, F., Roberto, C. and Scheffer, G. (2000). *Sur les Inégalités de Sobolev Logarithmiques. Panoramas et Synthèses [Panoramas and Syntheses]* **10**. Paris: Société Mathématique de France. With a preface by Dominique Bakry and Michel Ledoux. [MR1845806](#)
- [2] Bakry, D. and Émery, M. (1985). Diffusions hypercontractives. In *Séminaire de Probabilités XIX, 1983/84. Lecture Notes in Math.* **1123** 177–206. Berlin: Springer. [MR0889476](#)
- [3] Benamou, J.-D. and Brenier, Y. (1999). A numerical method for the optimal time-continuous mass transport problem and related problems. In *Monge Ampère Equation: Applications to Geometry and Optimization (Deerfield Beach, FL, 1997)*. *Contemp. Math.* **226** 1–11. Providence, RI: Amer. Math. Soc. [MR1660739](#)
- [4] Bobkov, S.G. and Chistyakov, G.P. (2015). Entropy power inequality for the Rényi entropy. *IEEE Trans. Inform. Theory* **61** 708–714. [MR3332742](#)

- [5] Brändén, P. (2011). Iterated sequences and the geometry of zeros. *J. Reine Angew. Math.* **658** 115–131. [MR2831515](#)
- [6] Carlen, E.A. and Gangbo, W. (2003). Constrained steepest descent in the 2-Wasserstein metric. *Ann. of Math.* (2) **157** 807–846. [MR1983782](#)
- [7] Cordero-Erausquin, D. (2002). Some applications of mass transport to Gaussian-type inequalities. *Arch. Ration. Mech. Anal.* **161** 257–269. [MR1894593](#)
- [8] Costa, M.H.M. (1985). A new entropy power inequality. *IEEE Trans. Inform. Theory* **31** 751–760. [MR0823597](#)
- [9] Gozlan, N., Roberto, C., Samson, P.-M. and Tetali, P. (2014). Displacement convexity of entropy and related inequalities on graphs. *Probab. Theory Related Fields* **160** 47–94. [MR3256809](#)
- [10] Hillion, E. (2012). Concavity of entropy along binomial convolutions. *Electron. Commun. Probab.* **17** Article no. 4. [MR2872573](#)
- [11] Hillion, E. and Johnson, O. (2016). Discrete versions of the transport equation and the Shepp–Olkin conjecture. *Ann. Probab.* **44** 276–306. [MR3456338](#)
- [12] Lott, J. and Villani, C. (2009). Ricci curvature for metric-measure spaces via optimal transport. *Ann. of Math.* (2) **169** 903–991. [MR2480619](#)
- [13] Mateev, P. (1978). The entropy of the multinomial distribution. *Teor. Veroyatnost. i Primenen.* **23** 196–198. [MR0490451](#)
- [14] Niculescu, C.P. (2000). A new look at Newton’s inequalities. *J. Inequal. Pure Appl. Math.* **1** Article No. 17. [MR1786404](#)
- [15] Rényi, A. (1957). A characterization of Poisson processes. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **1** 519–527. [MR0094861](#)
- [16] Rényi, A. (1961). On measures of entropy and information. In *Proc. 4th Berkeley Sympos. Math. Statist. and Probab.* **1** 547–561. Berkeley, CA: Univ. California Press. [MR0132570](#)
- [17] Shepp, L.A. and Olkin, I. (1981). Entropy of the sum of independent Bernoulli random variables and of the multinomial distribution. In *Contributions to Probability* 201–206. New York: Academic Press. [MR0618689](#)
- [18] Sturm, K.-T. (2006). On the geometry of metric measure spaces. I. *Acta Math.* **196** 65–131. [MR2237206](#)
- [19] Sturm, K.-T. (2006). On the geometry of metric measure spaces. II. *Acta Math.* **196** 133–177. [MR2237207](#)
- [20] Tsallis, C. (1988). Possible generalization of Boltzmann–Gibbs statistics. *J. Stat. Phys.* **52** 479–487. [MR0968597](#)
- [21] Villani, C. (2009). *Optimal Transport: Old and New. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **338**. Berlin: Springer. [MR2459454](#)
- [22] Wang, L. and Madiman, M. (2014). Beyond the entropy power inequality, via rearrangements. *IEEE Trans. Inform. Theory* **60** 5116–5137. [MR3252379](#)
- [23] Yu, Y. and Johnson, O.T. (2009). Concavity of entropy under thinning. In *Proceedings of the 2009 IEEE International Symposium on Information Theory* 144–148. New York: IEEE. DOI: [10.1109/ISIT.2009.5205880](#).

Received March 2015 and revised January 2016